THE ANALYST
FULL PAPER
www.rsc.org/analyst

Application of artificial neural networks to the classification of soils from São Paulo state using near-infrared spectroscopy

Paulo Henrique Fidêncio,[ab] Itziar Ruisánchez[a] and Ronei Jesus Poppi*[b]

[a] *Universitat Rovira i Virgilli, Departament de Química Analítica i Química Orgànica, Pça. Imperial Tàrraco 1, 43005, Tarragona, Spain*
[b] *Universidade Estadual de Campinas, Instituto de Química, CP 6154, 13083-970, Campinas, SP, Brazil. E-mail: ronei@iqm.unicamp.br; Fax: +55 19 37883023; Tel: +55 19 37883126*

Received 20th August 2001, Accepted 11th October 2001
First published as an Advance Article on the web 15th November 2001

This paper describes how artificial neural networks can be used to classify multivariate data. Two types of neural networks were applied: a counter propagation neural network (CP-ANN) and a radial basis function network (RBFN). These strategies were used to classify soil samples from different geographical regions in Brazil by means of their near-infrared (diffuse reflectance) spectra. The results were better with CP-ANN (classification error 8.6%) than with RBFN (classification error 11.0%).

1. Introduction

Nowadays, owing to the considerable advances in chemical instrumentation, it is easy and relatively inexpensive to obtain great quantities of data. The main problem is now how to process and interpret these data. Chemometric methods have been widely applied to evaluate and interpret data.[1] Moreover, multivariate data analysis searches for relationships among objects, among variables and among objects and variables. In this respect, considerable effort has been made to systemise data analysis. For example, Beebe *et al.*[2] proposed six steps that provide a recipe for the systematic evaluation of data: (1) examine, (2) pre-process as needed, (3) estimate the model, (4) examine the results/validate the model, (5) use the model for prediction and (6) validate the prediction.

According to this procedure, in order to solve a specific problem, the appropriate data should be collected, evaluated and used to obtain a mathematical model that should enable the problem to be solved. According to the specific problem, the model has to be validated.

The specific problem in this study was to identify the types of soils from four different geographical regions of São Paulo state, Brazil. Hence we are dealing with a classification problem. Soil samples (241) were collected from different regions and were analysed by near-infrared (NIR) spectroscopy. We were interested in obtaining information about soil types and in developing a fast, efficient strategy for classifying new, unknown samples. Soil classification is important because from the results it is possible to identify a specific region and to estimate several physical, chemical and morphological properties. Also, the soil classification defines the type of management and it is indicative of the cultivation to be implemented.

During the last decade, classical pattern recognition techniques such as principal component analysis, cluster analysis, canonical variate analysis, *etc.*, have been widely used in classification problems in all fields of chemistry.[3] Classical classification techniques such as SIMCA or KNN have also been widely used.[4–6] In addition to classical classification techniques, artificial neural networks (ANN) are now increasingly applied and have been reported in several publications.[7–11]

In this work, we evaluated the NIR data for the soils following the aforementioned six steps. Classical chemometric techniques such as SIMCA and KNN have been applied to the problem of determining soil types but have proved not to be successful, so we considered that they are not good options. Therefore, we used ANNs, specifically counter propagation and radial basis function neural networks, as classification techniques and the results are discussed here.

NIR spectroscopy for rapid, precise and non-destructive analysis has frequently been described as an alternative to conventional analytical techniques. The spectral features of the NIR spectral region are associated with the vibrational modes of the functional groups. These modes are combinations and overtones of the fundamentals in the IR region. Normally, the information contained in the spectra is difficult to interpret.

2. ANN theory

ANNs are well described in many publications. This section gives only a very brief description of the ANNs used here, namely Kohonen, counter-propagation and radial basis function learning strategies.

2.1 Kohonen artificial neural network

The Kohonen network[12,13] is a 'self-organising' system which is capable of solving unsupervised rather than supervised problems. It consists of only one layer of neurons in which each neuron contains as many weights as there are elements in the input vector $\boldsymbol{X}_s$; that is, it contains m variables. Therefore, the number of weights in each neuron coincides with the number of input variables. Neurons are, in the Kohonen ANN, ordered in a two-dimensional formation (see Fig. 1), dimensions are specified as $N \times N$. Fig. 1 shows a 21×21 Kohonen network with an input vector having 51 variables.

In the Kohonen approach, the neurons learn to pinpoint the location of the neuron in the network that is most 'similar' to the input vector $\boldsymbol{X}_S$. The Kohonen ANN automatically adapts itself in such a way that similar input objects excite most the neurons placed topologically close to each other. Weights are corrected after the input of each object using well described procedures[7,12–14] in such a way that not only is the weight of the neuron with the most 'distinguished' response corrected, but also the weights of the neurons in each neighbourhood. At the

 DOI: 10.1039/b107533k

end of the training the visualisation of the distribution of the objects can be seen on a table called a top-map.

An important characteristic is that because each level of weights in the Kohonen ANN always receives only one type of variable, at the end of the training in each weight level (51 levels in Fig. 1) a map showing the distribution values of the particular input variable is formed; this is called a contour map. By inspection of all weight levels and the corresponding maps, it can be found whether the distribution of the values of an *i*th variable corresponds to a group of objects found in the top-map of the Kohonen ANN, showing the importance of such variable in the formation and therefore characteristics of the group.[11]

2.2 Counter propagation artificial neural network

The counter-propagation ANN[15,16] is an adaptation of the Kohonen network and is able to solve supervised problems. In fact, it is a Kohonen ANN augmented by an output layer of neurons placed exactly below the neurons of the Kohonen layer (Fig. 2). The output layer of neurons, which has as many weight planes as the target vectors have responses, is corrected in a very similar way to the Kohonen layer.[7] However, there is an important difference. The most distinguished neuron in the output layer around which the corrections are made is not the one that is closest to the target, but the one exactly below the selected neuron in the Kohonen layer. The target vectors $\boldsymbol{T}_{\mathbf{S}}$ are binary vectors (t_{s1}, t_{s2}, …, t_{sn}) which have as many components as there are classes with which the object can be associated. Because one object $\boldsymbol{X}_{\mathbf{S}}$ can belong to only one class, only one of the components t_{si} of a given target $\boldsymbol{T}_{\mathbf{S}}$ is equal to one and all the other components t_{sj} *(ji)* are equal to zero. In our case, a four-dimensional target vector $\boldsymbol{T}_{\mathbf{S}}$, for a soil belonging to class No. 2, would be (0,1,0,0).

Once training is over, the Kohonen layer acts as a pointer device which for any $\boldsymbol{X}_{\mathbf{s}}$ determines the position in the network's output layer where the answer is stored. During the learning procedure, the answers (components of target vectors $\boldsymbol{T}_{\mathbf{S}}$) are distributed throughout the assembly of output neuron weights: target components t_i are distributed only in the *i*th level of output weights. Because each output component is distributed throughout the level of corresponding weights, each output neuron contains the answers for all classes even if its counterpart in the Kohonen layer above it was never excited during the training. The values of the weights in each output neuron *j* are distributed between 0 and 1 in such a way that the sum of all *n* output weights is equal to one.

The output weights w_{ji} of each *j*th output neuron are related to the probability that the object that excites this neuron is associated with class *i*. On the other hand, because all input weights carry values related to the variables, each two-dimensional plane (level) of input neurons can be regarded as a two-dimensional map of this variable. The overlap between the variable maps gives us an insight into how the variables are correlated among one other.[7] This is valid for weight maps either from the input or from the output layer. Both Kohonen and counter propagation ANN were performed using the NEURON and K-CTR set of programs.[17]

2.3 Radial basis function network (RBFN)

The RBFN is a type of neural network which is used to solve such problems as modelling and classification.[18] The RBFN can be considered as a two-layered network (see Fig. 3) with one hidden and one output layer. The input layer does not process the information, it only distributes the input variables to the hidden layer. Each neuron of the hidden layer represents a radial function and the number of radial functions depends on the problem to be solved.

The radial functions which are most used are Gaussian functions.[19,20] They are characterised by two parameters: the centroids, represented by c_j, and the width, represented by σ_j (see Fig. 3). The outputs from the radial functions are fully connected to the neurons of the output layer by the strength or weight coefficients w_{kj}. Finally, the response of each output neuron is calculated by a linear least-squares regression of its inputs, that is, the output of the hidden layer. For a fixed σ_j value, the learning procedure means finding the optimum c_j and w_{kj} in order to obtain minimum RMS values, taking into account the target vector. The target vector is exactly the same as that described in the counter propagation neural network. The mathematical expressions used to adjust c_j and w_{kj} are well

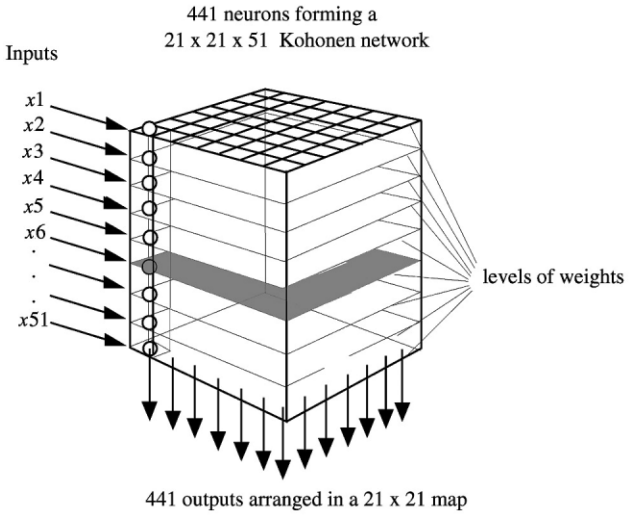

Fig. 1 Kohonen ANN architecture used in the present study and represented as a 21 × 21 block of neurons. Each neuron is repesented as a column of 51 weights. The level of weights which handles the sixth input variable x_6 is shown in grey. The sixth weight in the neuron at the position (1,1) is also shaded.

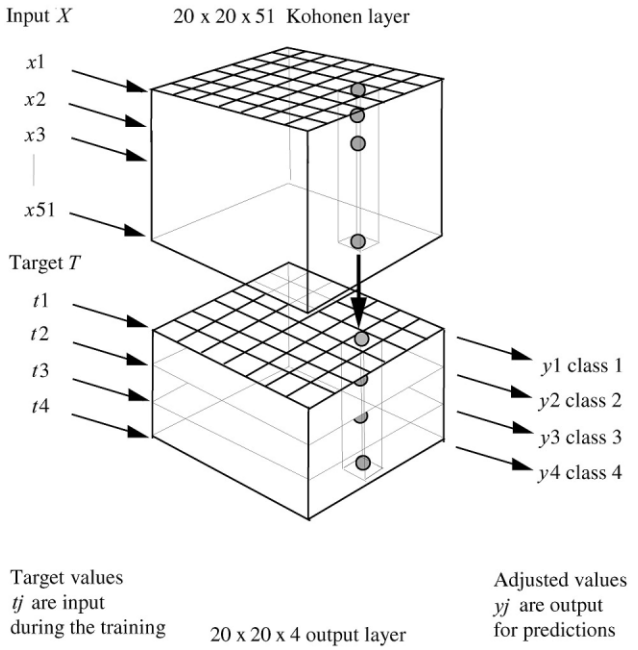

Fig. 2 Counter propagation ANN consists of two layers of neurons: the Kohonen and the output layers. In the Kohonen layer to which the objects $\boldsymbol{X}_{\mathbf{s}}$ are input, the central neuron is selected and then corrections of weights are made around its position (bold arrow) in the Kohonen and in the output layer.

described in the literature.[18] All calculations were made with a program implemented in the Matlab environment.[21]

3 Experimental

3.1 Samples

The soil samples were collected at different localities in the state of São Paulo: Garça, Itararé, Pariquera-Açu, Assis, Araçatuba, Americana, Marília, Votoporanga, Adamantina, Jaú, Campinas, Piracicaba and Cosmópolis. They were analysed by the soil bank of the Instituto Agroñomico de Campinas (IAC)—Brazil. The samples were of different colours and were collected at depths between 0–10 and 80–100 cm.

The soil samples were assigned by IAC to four pre-defined types or classes. The first class (44 samples) is called oxisol and is a dark red latosol; the second class (50 samples) is also called oxisol but is a dusky red latosol; the third class (48 samples) is also called oxisol and is a red–yellow latosol; the fourth class (49 samples) is called ultisol and is a red–yellow podzolic (see Table 1). Oxisol soils are strongly acidic, have low capacities for cation exchange and have an Fe_2O_3 content between 18 and 36%. Utisol soils are neither acidic nor alkaline, have a good capacity for cation exchange and a low content of aluminium.[22] For this study, we had 191 samples whose soil type was known and 50 samples whose soil type was unknown.

3.2 Near-infrared analyses

Initially, the soil samples were gently crushed to pass through a <0.177 mm sieve and dried at 40 °C for 24 h. The diffuse reflectance spectra of the powdered soils were recorded on a CARY 5G UV/VIS/NIR spectrophotometer, set in the NIR region (1000–2500 nm). A constant spectral resolution value of 4.4 nm was used, which gives a 241 × 339 (reflectance values) matrix. The spectra in reflectance (R) were converted to log(1/R) units. The soil reflectance was reported relative to a $BaSO_4$ standard spectrum. Examples of the different diffuse reflectance spectra of the four classes are shown in Fig. 4.

4. Chemometric procedure

4.1 Data pre-processing

An important stage in the data analyses is pre-processing. The importance lies in the fact that each pre-treatment has a significative influence on the modelling step, generally when the data are complex, as is the case with soil spectra. Finding the optimum pre-processing is very time consuming, as many types of pre-processing have to be tested and evaluated according to the model obtained. All of the 241 spectra obtained are defined by their log(1/R) values, equivalent to the absorbance values, and, as has been stated above, this resulted in 339 variables. This is a large number of variables, so to reduce this number, one of every four values was chosen and the remaining three discarded. The first 34 absorbance values were also eliminated, as it was found that they do not contribute relevant information to the classification problem (see below). After the spectra had been reduced, the data were pre-processed with multiplicative scatter correction (MSC).[23] From these data, matrices of 241 samples × 51 variables were obtained.

4.2 Data exploration

In our first attempt to explore the data set, the diffuse reflectance spectra were plotted in fewer dimensions in the PCA space. PCA is a well established method and it is powerful as a display technique but its results are difficult to interpret when there are numerous samples or when the percentage of explained variance is not high enough. Fig. 5 shows the PCA plot, which explains 99.9% of the variance, obtained for 191 NIR soil samples with the pre-treatment specified above. It can be clearly seen that three of the classes overlap, so is not possible to differentiate among them. Class 2, however, is separate from the rest. It can also be seen that classes are ordered along the first PC, which explains the maximum variance.

The 191 spectra were mapped from the 51-dimensional variable space on to a two-dimensional space by means of Kohonen mapping. To ensure that most of the 191 samples would excite different neurons, an ANN architecture of 21 × 21 × (51) was used. All the neurons in this architecture had 51 weights resulting from the 51 transformed input variables. Fig.

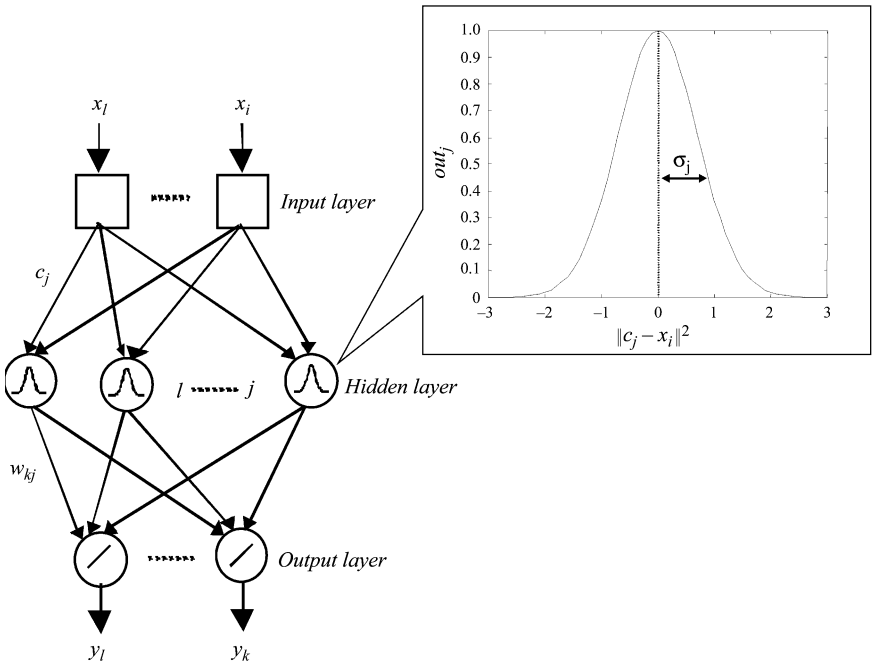

Fig. 3 Radial basis function network architecture containing one hidden layer with j neurons and the output layer with k neurons.

6 shows that with Kohonen mapping it was possible to differentiate among all four classes, although class 4 was split in two. This is consistent with the spectra shown in Fig. 5, which show two types of spectra for the same type of soil. As in the PCA plot, class 2 is separate from the other classes. Class 2 soils are dusky red latosols, whose chemical characteristics are different from those of the other three classes.[22] Soils in classes 1, 3 and 4 are closer together. Characteristic soils from classes 3 and 4 are yellow–red whereas those from class 1 are dark red. The red colour might be an important or significant difference between the soils in classes 1, 3 and 4 and those in class 2.

4.3 Training and test set selection

Once we had found that is possible to differentiate among all four classes by means of the Kohonen network and in order to assign unknown samples to their geographical origin, a classification model was developed using an artificial neural network: counter-propagation and RBF. To establish a classification rule, the data matrix had to be divided into a training set and a test set. Several approaches are available for selecting training and test sets. There are some key points that have to be taken into account as both sets have to be representative of the whole data set and all possible sources of variation that can be encountered later must be included in the training set. Among the several approaches, the simplest is the random selection but it is open to the possibility that some sources of variations will be lost. Several algorithms such as Kennard–Stone,[24] D-optimal, *etc*., have been proposed, mostly for calibration problems. Finally, whenever dealing with cluster data, clustering or representation techniques such as PCA, Kohonen, *etc*., are recommended.[25,26]

In our particular problem, the selection was based on the Kohonen map shown in Fig. 6. and took into account that samples from each class have to be representative and homogeneously distributed. Hence, from each neuron excited by more than one object, only one was taken, never two or more from the same neurons. The reason for this is that similar objects, which excite the same neuron, do not contribute new information to the modelling. For each class, the relative position of the objects in the Kohonen plot was studied and the criterion was to select the maximum number of objects that define the borders of each class and some from the centre. So the training set was made up of 63 objects (which are circled in Fig. 6) and the test set was made up of the remaining 128 objects.

Table 1 Identification of class

Class	Identification	Sample numbers
1 (○)	Dark red latosol	1–44
2 (+)	Dusky red latosol	45–94
3 (*)	Red–yellow latosol	95–142
4 (Δ)	Red–yellow podzolic	143–191

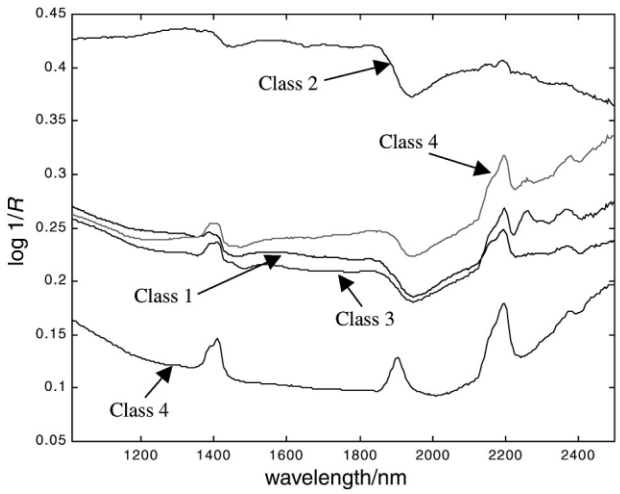

Fig. 4 Examples of diffuse reflectance NIR spectra for the four predefined classes: dark red latosol (class 1), dusky red latosol (class 2), red–yellow latosol (class 3) and red–yellow podzolic (class 4).

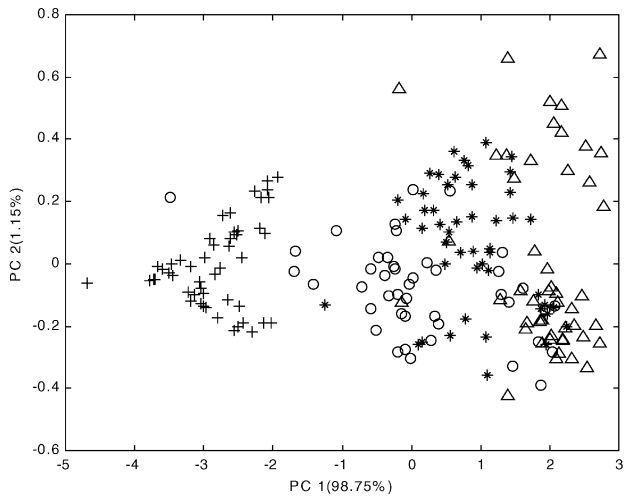

Fig. 5 PCA plot obtained for the 191 NIR spectra of soil samples.

4.4 Network architecture

The CP-ANN[7,8] architecture for building the classification model with the selected training set consists of 20 × 20 × (51 + 4). Therefore, we had 51 weights in the Kohonen layer (input variables) and 4 weights in the output layer (four classes) (Fig. 2). The maximum number of iterations in the training process was 5000 and the function used to correct the weights on neighbouring neurons was the constant function. This function provides better results than the triangular and Mexican hat functions.

Basically, in RBFN it is necessary to optimise and define two parameters, the number of neurons in the hidden layer and the width of the radial basis functions (σ). The optimum values of these parameters were chosen to provide minimum training and validation set RMS errors. As we are dealing with an analytical classification problem, the type of classification error must also be considered. We tested 1–60 hidden neurons and σ = 0.1–1. The best results were obtained using 35 radial neurons in the hidden layer, which is a rather large number of neurons, with σ = 0.1. Hence the final architecture of the RBF network applied can be written as 51 × 35 × 4.

5. Results and discussion

5.1 CP-ANN classification

As has already been stated, the counter propagation ANN used for prediction was obtained by adding a four-level output layer to a 20 × 20 Kohonen ANN. Each weight level of the output layer corresponds to one of the four classes to be predicted.

Table 2 shows the results of classifying the 128 test samples. For class 1, 86% of the samples were correctly assigned, and three samples (numbers 9, 29 and 41) were incorrectly assigned as belonging to class 3. One sample (number 27) was not assigned to any of the four classes. All samples belonging to class 2 were correctly assigned, so we had a 100% correct classification. Class 2 was the most different class, as indicated by the PCA plot (Fig. 5) and the Kohonen map (Fig. 6). For

class 3, 83% of the samples were correctly assigned. However, four samples were incorrectly assigned: two of them (numbers 138 and 142) were assigned to class 1 and the other two (numbers 128 and 139) were assigned to class 4. One sample (number 109) was not assigned to any of the four classes. Finally, for class 4, 88% were correctly classified and four samples were incorrectly assigned: one (number 144) was assigned to class 1 and three (numbers 146, 155 and 165) to class 3. If we look at the total classification error for the 128 test set samples, only 11 samples were incorrectly assigned. This is an error of 8.60%. Two of the samples were not assigned to any of the four classes, an error of 1.56%.

5.2 RBFN classification

As has been stated above, the number of neurons in the hidden layer and the value of σ were optimised to 35 and 0.1, respectively, with the final architecture of the RBF network being 51 × 35 × 4.

Table 3 shows the results of classifying the 128 test samples. For class 1, 72% of the samples were correctly assigned. However, four samples (numbers 14, 25, 29 and 31) were assigned to class 3 and four samples (numbers 27, 28, 30 and 36) were not assigned at all. For class 2, 88% of the samples were correctly assigned. Two samples, however, were incorrectly assigned; one (number 51) was assigned to class 2 and the other (number 64) to class 3. Two samples (numbers 53 and 91) were not assigned to any of the pre-established classes. These results are worse than those obtained with CP-ANN,in which classification was 100%. For class 3, 73% of the samples were correctly assigned. Seven samples were incorrectly assigned: five (numbers 136, 137, 138, 141 and 142) to class 1 and two (numbers 126 and 139) to class 4. One sample (number 135) was not assigned to any of the four pre-defined classes. Finally, for class 4, 91% of the samples were correctly assigned. One sample (number 144) was incorrectly assigned to class 3 and two samples (numbers 155 and 176) were not assigned to any of the four classes. In this case, the results obtained with RBF are better than those obtained with CP-ANN, where classification was only 88%. If we look at the classification for the 128 test set samples, 14 samples were incorrectly assigned, an error of 10.94%. Moreover, when RBFN was used, nine samples were not assigned, 7.03% of the total number of samples subjected to the test.

5.3 Classification of unknown samples

The classification models developed by CP-ANN and RBFN were used to classify 50 unknown samples, of which we had no *a priori* knowledge of their soil type according to the IAC information. The only information provided was that they were collected in the same area as those used for the training and test

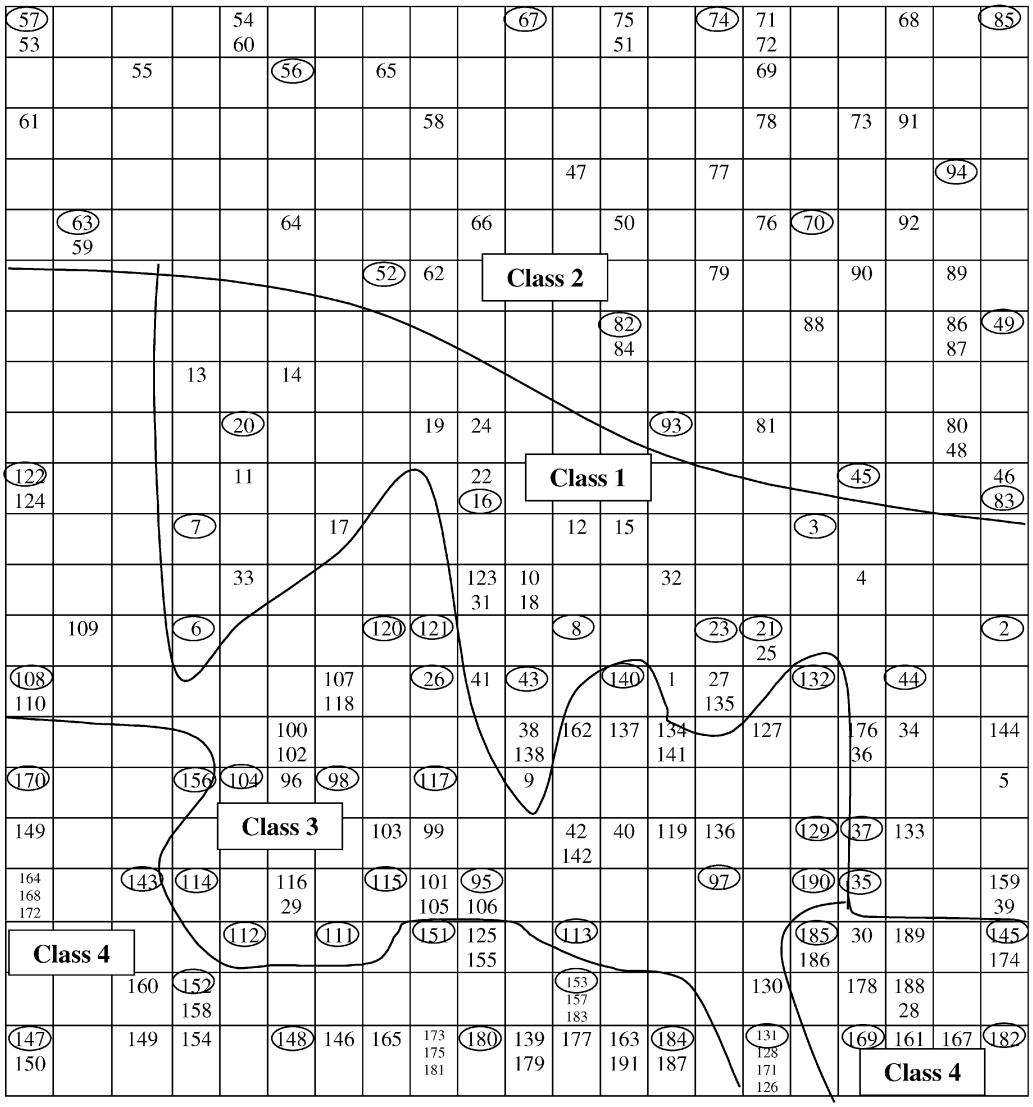

Fig. 6 Kohonen map for the 191 soil samples used in the selection of the training and test data sets. Class 2 is the most separated. 63 samples were selected for the training set (they are circled) and 128 samples were used for the test set.

sets. Working with the trained set with CP-ANN (see Table 4), 45 samples were assigned to one of the four classes and 5 were not assigned to any of the pre-established classes. Not classified means that the objects fell in the border between two pre-defined groups on the top-map and therefore their assignment to one of the groups or classes is not possible. Using RBFN it was not possible to classify any soil sample in the pre-defined classes (see Table 4).

5.4 Variable information contour maps

As mentioned above, Kohonen ANN yields a contour map for each input variable. In our case, these variables are the absorbances at different wavelengths. These contour maps were used to evaluate how important the variables are at separating and therefore defining each of the four classes. The first 34 variables were eliminated from the initial 85 reduced variables, as it had been shown that they did not contribute any significant information. The variables were therefore reduced to 51, which were used to build the final classification models. The importance of the 51 variables for the shape and position of the four classes in the Kohonen map was also analysed. Fig. 7 shows an overlap of the output layer (top-map) obtained after training with the 63 training samples and the contour maps of the most significant variables. It can be seen that class 1 is characterised by variables 49, 50 and 51 (solid line), which correspond to absorbance values at 2453, 2471 and 2488 nm, respectively (this region is characteristic of the absorption of phenolic O–H and aliphatic C–H). Class 2 is characterised by variables 32, 33, 34 and 35 (dashed line), which correspond to absorbance values at 2154, 2172, 2189 and 2207 nm, respectively (this region is characteristic of the absorption of phenolic O–H, amine N–H and amide N–H) and variables 43, 44, 45 and 46, which correspond to absorbance values at 2348, 2365, 2383 and 2400 nm, respectively (this region is characteristic of the absorption of aliphatic C–H). Class 4 is characterised by variables 16, 17, 18, 19, 20, 21, 22 and 23 (dotted line), which correspond to absorbance values at 1872, 1890, 1908, 1925, 1943, 1960, 1978 and 1996 nm, respectively (this region is characteristic of the absorption of phenolic O–H, amide N–H and aliphatic C–H). Finally, no specific variable was observed as being important for the definition of class 3. Fig. 7 shows an overlap of the input variables contour maps and it can be seen that the combination allows differentiation among all four classes.

Table 2 Overall prediction results for the validation set objects using CP-ANN

	Predicted class					
Real class	1	2	3	4	Not classified	Correct (%)
1	25	—	3	—	1	86
2	—	35	—	—	—	100
3	2	—	25	2	1	83
4	1	—	3	30	—	88

Table 3 Overall prediction results for the validation set objects using RBFN

	Predicted class					
Real class	1	2	3	4	Not classified	Correct (%)
1	21	—	4	—	4	72
2	1	31	1	—	2	88
3	5	—	22	2	1	73
4	—	—	1	31	2	91

Table 4 Overall prediction results for the test set objects using CP-ANN and RBFN

	Predicted class				
Method	1	2	3	4	Not classified
CP-ANN	4	—	6	35	5
RBFN	Not possible				

Conclusions

The main aim of this study was to provide an insight into multivariate soil NIR data. ANN was shown to give the clearest information, Kohonen ANN enabled all four classes to be differentiated according to the distribution made by the IAC whereas PCA only separated one class (class 2) from the rest. In the Kohonen map, the classes can be better seen as individual groups and class 2 is more separate from the other classes. The contour maps obtained by means of Kohonen ANN overlapped with the top-map providing additional information about how each input variable influenced the separation of the four classes of soils.

Classification was possible by means of CP-ANN and RBFN although the results were better with CP-ANN. It must be emphasised that several samples were incorrectly assigned or not assigned at all as they fell on the borders between their own class and another one. In these cases, although there is no clear classification, some information can be extracted. In the present study these samples were considered as incorrectly assigned.

Acknowledgements

The authors thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for a fellowship to P. H. Fidêncio (Proc. 97/04062-6) and the chemometrics group of the

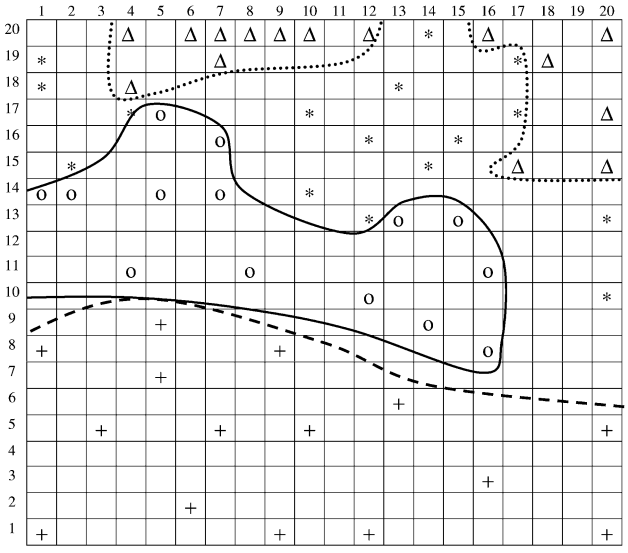

Fig. 7 Overlap of the 20 × 20 Kohonen output layer (top-map) showing the object distribution and the contour maps corresponding to the input variables. The most significant input variables are represented as a line that defined the borders of each class. Class 1 (○, solid line), class 2 (+, dashed line), class 3 (*) and class 4 (Δ, dotted line).

Universitat Rovira i Virgili for an agreeable period of collaboration.

References

1 B. G. M. Vandeginste, D. L. Massart, L. M. C. Buydens, S. de Jong, P. J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics, Part B*, Elsevier, Amsterdam, 1998.
2 K. R. Beebe, R. J. Pell and M. B. Seasholtz, *Chemometrics: a Practical Guide*, Wiley, New York, 1998.
3 M. Defernez and E. K. Kemsley, *Trends Anal. Chem.*, 1997, **16**, 216.
4 M. Núñez, R. M. Peña, C. Herrero and S. Garcia-Martin, *Analusis*, 2000, **28**, 432.
5 P. M. Padin, R. M. Peña, S. Garcia, R. Iglesias, S. Barro and C. Herrero, *Analyst*, 2001, **126**, 97.
6 G. Downey, J. McElhinney and T. Fearn, *Appl. Spectrosc.*, 2000, **54**, 894.
7 J. Zupan and J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, Wiley-VCH, New York, 1999.
8 J. Zupan, M. Novic and I. Ruisánchez, *Chemom. Intell. Lab. Syst.*, 1997, **38**, 1.
9 W. Wu and D. L. Massart, *Chemom. Intell. Lab. Syst.*, 1996, **35**, 127.
10 J. A. Remolà, J. Lozano, I. Ruisanchez, M. S. Larrechi, F. X. Rius and J. Zupan, *Trends Anal. Chem.*, 1996, **15**, 137.
11 I. Ruisánchez, P. Potokar, J. Zupan and V. Smolej, *J. Chem. Inf. Comput. Sci.*, 1996, **36**, 214.
12 T. Kohonen, *Self-Organisation and Associative Memory*, Springer, Berlin, 1988.
13 T. Kohonen, *Proc. IEEE*, 1990, **78**, 1464.
14 T. Kohonen, *Biol. Cybern.*, 1982, **43**, 59.
15 R. Hecht-Nielsen, *Appl. Opt.*, 1987, **26**, 4979.
16 R. Hecht-Nielsen, *Neural Networks*, 1988, **1**, 131.
17 J. Zupan, *Program K-CTR for Kohonen and Counter Propagation ANNs*, Report-KI DN-1452, 1994.
18 A. Pulido, I. Ruisánchez and F. X. Rius, *Anal. Chim. Acta*, 1999, **388**, 273.
19 E. P. P. A. Derks, M. S. Sánchez and L. M. C. Buydens, *Chemom. Intell. Lab. Syst.*, 1995, **28**, 49.
20 B. Walczak and D. L. Massart, *Anal. Chim. Acta*, 1996, **331**, 177.
21 *MATLAB 5.3*, MathWorks, Natick, MA, 1999.
22 *Brazilian System for the Classification of Soils*, Embrapa, Brasilia, 1999.
23 P. Geladi, D. MacDougall and H. Martens, *Appl. Spectrosc.*, 1985, **39**, 491.
24 R. W. Kennard and L. A. Stone, *Technometrics*, 1969, **11**, 137.
25 N. Majcen, K. Rajer-Kanduc, M. Novic and J. Zupan, *Anal. Chem.*, 1995, **67**, 2154.
26 J. Zupan, M. Novic and I. Ruisánchez, *Chemom. Intell. Lab. Syst.*, 1997, **38**, 1.